



Wavelet Shrinkage: Unification of Basic Thresholding Functions and Thresholds

Abdourrahmane Atto, Dominique Pastor, Grégoire Mercier

► To cite this version:

Abdourrahmane Atto, Dominique Pastor, Grégoire Mercier. Wavelet Shrinkage: Unification of Basic Thresholding Functions and Thresholds. *Signal, Image and Video Processing*, 2011, 5 (1), pp.11-28. 10.1007/s11760-009-0139-y . hal-00732360

HAL Id: hal-00732360

<https://hal.science/hal-00732360>

Submitted on 14 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Wavelet Shrinkage: Unification of Basic Thresholding Functions and Thresholds

Abdourrahmane M. Atto, ^{*} Dominique Pastor, [†] Grégoire Mercier [‡]

Abstract

This work addresses the unification of some basic functions and thresholds used in non-parametric estimation of signals by shrinkage in the wavelet domain. The Soft and Hard thresholding functions are presented as degenerate *smooth sigmoid based shrinkage* functions. The shrinkage achieved by this new family of sigmoid based functions is then shown to be equivalent to a regularisation of wavelet coefficients associated with a class of penalty functions. Some sigmoid based penalty functions are calculated, and their properties are discussed. The unification also concerns the universal and the minimax thresholds used to calibrate standard Soft and Hard thresholding functions: these thresholds pertain to a wide class of thresholds, called the detection thresholds. These thresholds depend on two parameters describing the sparsity degree for the wavelet representation of a signal. It is also shown that the non-degenerate sigmoid shrinkage adjusted with the new detection thresholds is as performant as the best up-to-date parametric and computationally expensive method. This justifies the relevance of sigmoid shrinkage for noise reduction in large databases or large size images.

^{*}TELECOM Bretagne, am.atto@telecom-bretagne.eu

[†]TELECOM Bretagne, dominique.pastor@telecom-bretagne.eu

[‡]TELECOM Bretagne, gregoire.mercier@telecom-bretagne.eu

keywords: Non-parametric Estimation, Wavelets, Shrinkage function, Penalty function, Detection thresholds.

1 Introduction

The soft and hard shrinkage (thresholding) functions are basic functions widely used for estimating a signal *via* projection in the wavelet domain. The soft and hard shrinkages [1] involve forcing to zero the coefficients with amplitudes lower than the selected threshold, and preserving (hard) or shrinking (soft) any coefficient, with amplitude above this threshold, by a value that equals the threshold height. Threshold selection for calibrating soft and hard thresholding functions has also been addressed by Donoho and Johnstone in [1]. These authors proposed the use of the universal and minimax thresholds: the estimation by soft or hard thresholding with any of these thresholds yields near-optimal risk in the sense that, asymptotically, the estimator achieves within a factor of $2 \log N$ of the ideal risk, which is the risk achieved with the aid of an oracle (see Donoho and Johnstone's paper for further details).

However, in practice, the hard and the soft WaveShrink estimators present drawbacks such as an important variance, when using hard thresholding, or a large bias, when using soft thresholding [2]. Many suggestions have been made in order to improve the performance of these WaveShrink estimators. Some of them relate to the choice of the threshold ([3], [4]), and others address the choice of the shrinkage (parametric Bayesian shrinkage: [5, 6, 7, 8, 9, 10]; non-parametric shrinkage functions: [11], [12], [13], [14]; among others).

The different contributions proposed in the literature and aiming at improving the denoising performance have resulted in a huge number of wavelet based methods for image denoising. In addition, there exist many ways to improve a given method (using suitable wavelet transform, adding intra-inter-scale predictors, exploiting redundancy, combining several methods, and so

forth). Actually, the most efficient denoising algorithms such as [8] and [13] combine several of these techniques (interscale predictors, Gaussian smoothing, laborious parameterizations) and thus, they lose the simplicity (single function with explicit close form) and the portability (using different wavelet transforms without additional computations) of basic shrinkage functions.

However, note that processing large size signals and images requires computationally fast techniques and processing large databases requires portability of the method. We are thus interested in efficient denoising by wavelet shrinkage when the shrinkage function has an explicit close form, without any additional *a priori* consideration such as interscale predictor.

In this respect, the present work revisits the concept of shrinkage function by addressing the consequences of two recent results: the *Smooth Sigmoid Based Shrinkage* (SSBS) functions of [14] and the detection thresholds of [4].

The SSBS functions are smooth functions and they allow for a flexible control of the shrinkage through parameters that model the attenuation imposed to small, median and large data. This makes it possible to correct the main drawbacks of the soft and hard shrinkage functions. In contrast to the “sum of derivative of Gaussian” parameterization of [13], the SSBS functions are defined by an explicit close form so that we can first adapt their shape according to the noise level and the expected denoising level; in addition these functions can be used for any wavelet transform (orthogonal, redundant, multi-wavelets, complex wavelets, among others) without additional computation, which is not the case for the methods such as the SURELET of [13] and the BLS-GSM of [8].

The detection thresholds are synthesized by considering a risk function chosen to be the probability of error for deciding that a coefficient is significant or not. They depend on two parameters that can be used to bound the sparsity degree of the wavelet representation [15]. These thresholds are optimal in the sense that, for a certain class of signals, including sparse signals, they lead to the same upper bound for the probability of error than the Bayes test with

minimal probability of error among all possible tests [4]. It is shown below that the standard minimax and universal thresholds are detection thresholds corresponding to different degrees of sparsity.

Summarizing, the present paper extends the results established in [14] by providing **1)** the general description of the SSBS parameters, that is, the relation that allows for computing the SSBS attenuation degree with respect to the SSBS threshold and the asymptotic attenuation parameters (this has been addressed in [14] only for the case of vanishing asymptotic attenuation); **2)** the penalty functions associated with the SSBS functions in a regularization problem; **3)** the combination between the SSBS functions and the detection thresholds defined in [4]; **4)** experimental results emphasizing the relevance of SSBS functions, combined with detection thresholds, in image denoising of synthetic and real noisy data.

The organization of the paper is as follows. Section 2 briefly describes the SSBS functions and provides their parameter interpretation. This section also highlights that the hard and soft thresholding functions can be seen as limit SSBS functions. Section 3 provides the characterization of the SSBS functions in a regularization problem by computing the SSBS penalty functions. Section 4 addresses the properties of the detection thresholds. This section also discusses the selection of appropriate detection thresholds with respect to the wavelet decomposition properties of some signals such as smooth and piecewise regular ones. Section 5 presents experimental tests aimed at assessing the denoising quality achieved by using the SSBS functions combined with detection thresholds. Finally, section 6 concludes this work.

2 Unification of Basic Thresholding functions

In what follows, we consider standard wavelet-based estimation procedures for discrete time signals with dyadic sample sizes, equally spaced and fixed sample points. We use the standard

model

$$c_i = d_i + \epsilon_i, \quad i = 1, 2, \dots, N, \quad (1)$$

where $\mathbf{c} = \{c_i\}_{1 \leq i \leq N}$ is the orthonormal wavelet transform of a noisy observation, $\mathbf{d} = \{d_i\}_{1 \leq i \leq N}$ is a sparse vector representing the wavelet coefficients of the unknown deterministic signal and noise $\boldsymbol{\epsilon} = \{\epsilon_i\}_{1 \leq i \leq N}$ is such that the random variables $\{\epsilon_i\}_{1 \leq i \leq N}$ are independent and identically distributed (iid), Gaussian, with null mean and variance σ^2 . In short, $\epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ for $i = 1, 2, \dots, N$. Vector \mathbf{d} is supposed to be sparse, meaning that the wavelet basis concentrates a large proportion of the energy of the signal in a small number of coefficients with large amplitudes. This heuristic notion of sparsity ranges from *strong* to *weak* sparsity.

By strong sparsity, we mean that the energy of the signal is “almost entirely” contained in a small number of coefficients with large amplitudes (see the example given in figure 1). In this case, almost all the coefficients described as “small” are in fact quasi-null or with very small amplitudes, and so, do not contain significant information on the signal. For this reason, thresholding rules like hard or soft thresholding are proved to be the relevant strategies for estimating the signal (see [1]).

In the case of a representation which is not strongly sparse, it may often be useful to process small coefficients. In fact, wavelet representations of natural images fail to be sparse enough (see the example of figure 2): textures, contours are characterized by many small coefficients, and forcing all the small coefficients to zero may result in over-smoothing these image characteristics and a loss of significant information when the threshold height is large. Thus, in such a case of weakly sparse representation, it may be preferable to consider, not a thresholding function, but a shrinkage function that performs a penalized shrinkage without systematically forcing to zero the small coefficients [14]. The family of Smooth Sigmoid-Based Shrinkage (SSBS) functions (introduced in [14]) are shrinkage functions of that kind.

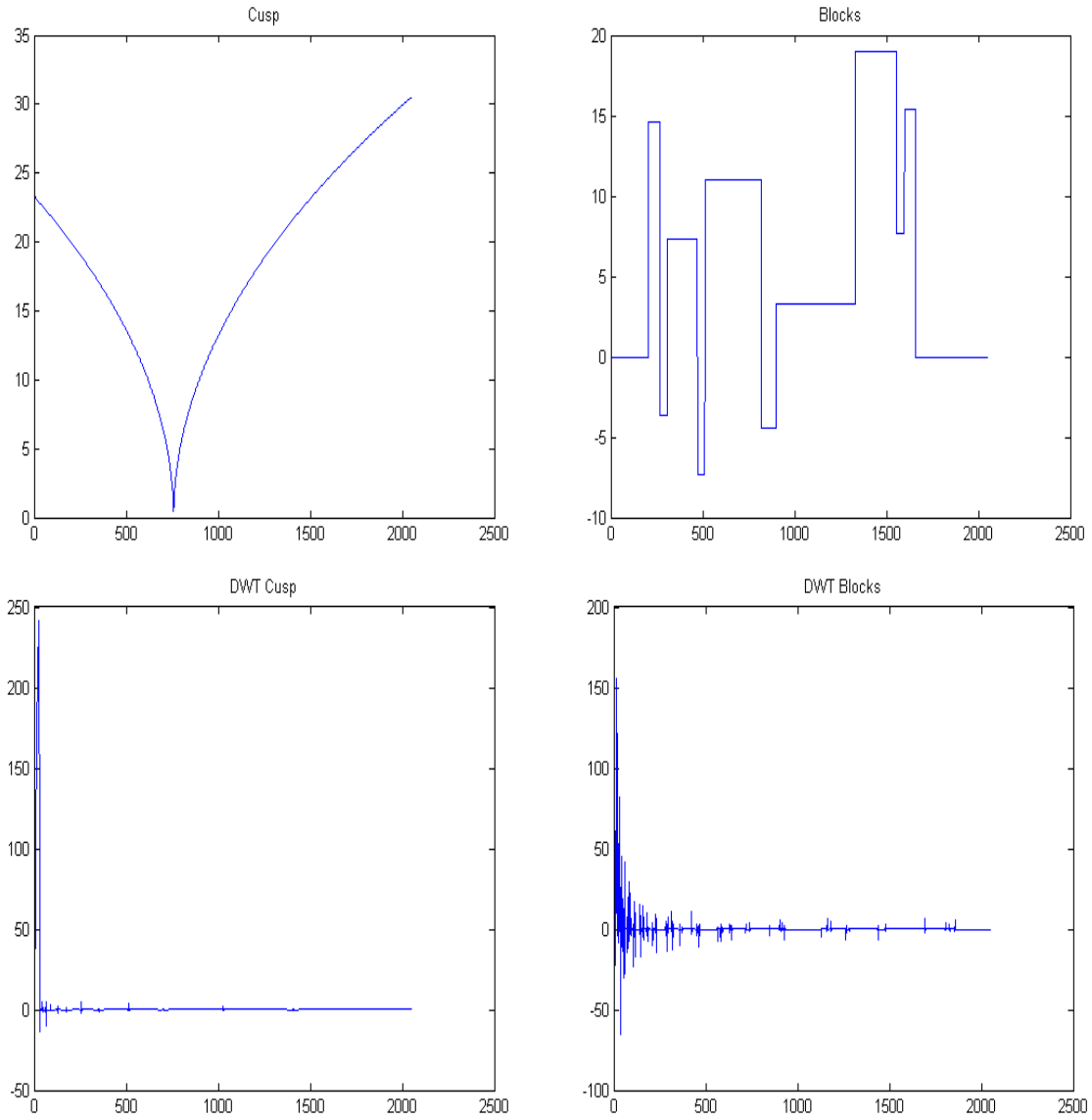


Figure 1: The discrete wavelet transform (DWT) of the ‘Cusp’ signal is strongly sparse. Almost all the energy of the ‘Cusp’ wavelet coefficients is concentrated in a very few number of large coefficients. In contrast, the sparsity of the wavelet representation of the ‘Blocks’ signal must be understood in the weak sense: this representation admits many small, but significant coefficients, because these coefficients characterize the singularities of the ‘Blocks’ signal. The Symlet wavelet of order 8 and 4 decomposition levels are used for the wavelet representation.

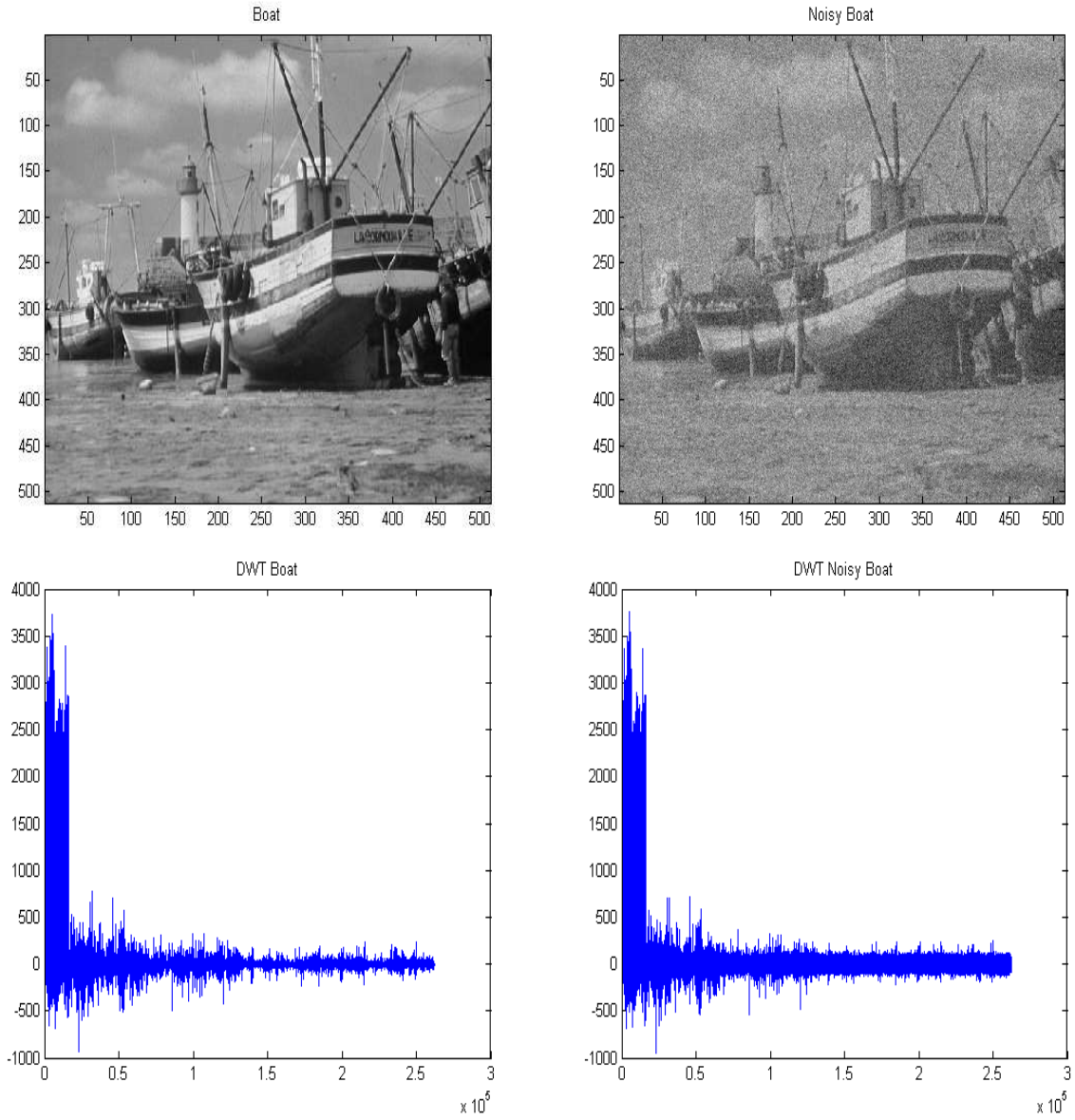


Figure 2: The wavelet representation (DWT) of natural images such as the 8-bit grayscale ‘Boat’ image fails to be sparse enough. The Symlet wavelet or order 8 and 4 decomposition levels are used for the wavelet representation. The noise standard deviation is 35.

Remark 1

The model of Eq. (1) (in the wavelet domain) is justified when the input signal (in the temporal or spatial domain) is corrupted by Additive White Gaussian Noise (AWGN) and the wavelet transform used is orthonormal. In the case where additive noise is either colored or not Gaussian, this model remains approximately valid in the following sense. The wavelet transform has interesting asymptotic statistical properties. In fact, the coefficients returned by the wavelet transform tend to be iid Gaussian when the decomposition level is large enough for stationary random processes [16, 17, 18, 19, 20] and some non-stationary random processes such as fractional Brownian motions or fractionally differenced processes [21, 22, 23, 24, 25]. Thus, for a large class of random noises, one can expect that the wavelet coefficients are quasi-decorrelated and approximately Gaussian distributed for large resolution levels. Note that the above property might not be satisfied at the first resolution levels for strongly correlated processes and the wavelet coefficients may thus remain strongly correlated at resolution levels that are not large enough. In this case, a solution would be to use the coefficients, provided by a full wavelet packet decomposition, at large enough resolution levels. Indeed, the wavelet packet decomposition also returns coefficients that tend to be iid Gaussian at every node of sufficiently large resolution levels. However, the convergence is more intricate than in the case of the wavelet transform, because the wavelet packet decomposition filters play an important role in this convergence. In fact, the asymptotic decorrelation and Gaussianity at large enough decomposition levels have been established for decomposition filters that are sufficiently regular. For filters with smaller regularity, the convergence to sequences of iid Gaussian wavelet packet coefficients remains an open issue. With respect to the foregoing, and since, for natural images, which are piecewise regular rather than smooth, sparse wavelet representations are better achieved by using small regularities, only wavelet transforms are considered below. In case of correlated noise, the threshold heights must be *level-dependent* because the noise wavelet coefficients have different

standard deviations depending on the resolution level.

The SSBS functions are the family of real-valued functions defined by [14]:

$$\delta_{t,\tau,\lambda}(x) = \frac{\text{sgn}(x)(|x| - t)_+}{1 + e^{-\tau(|x| - \lambda)}}, \quad (2)$$

for $x \in \mathbb{R}$, $(t, \tau, \lambda) \in \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+$, where $\text{sgn}(x) = 1$ (resp. -1) if $x \geq 0$ (resp. $x < 0$), and $(x)_+ = x$ (resp. 0) if $x \geq 0$ (resp. $x < 0$). Each $\delta_{t,\tau,\lambda}$ is the product of the soft thresholding function with a sigmoid-like function. As such, the function $\delta_{t,\tau,\lambda}$ is called a Smooth Sigmoid-Based Shrinkage (SSBS) function. The soft and hard thresholding functions can be regarded as SSBS functions for degenerate values of the parameter τ (see Appendix A). In the rest of the paper, the soft and hard thresholding functions will be referred as degenerate SSBSs.

The following addresses the role of the SSBS parameters. First, note that the parameter t controls the attenuation imposed to data with large amplitudes (see figure 3). Thus, it will be called the *asymptotic attenuation* parameter. In the rest of the paper, we assume that $\lambda \geq t$. In

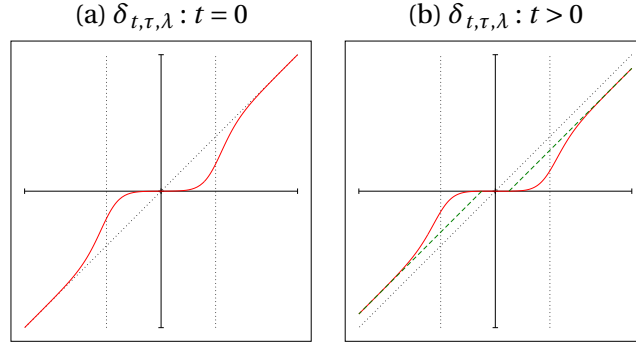


Figure 3: Graphs of $\delta_{t,\tau,\lambda}$ for $t = 0$ and $t \neq 0$. The dotted lines represent the values $\pm\lambda$.

fact, if $\lambda < t$, then $\delta_{t,\tau,\lambda}$ behaves as the soft thresholding function $\text{sgn}(x)(|x| - t)_+$ when $x \leq t$ and x is large. Unfortunately, soft thresholding is known to over-smooth the estimate when t is either the universal or the minimax threshold.

Second, the parameter λ will be described as the SSBS *threshold* since it acts as a threshold: $\delta_{0,\infty,\lambda}$ is a hard thresholding function with threshold height λ .

Finally, after a re-parametrization of the SSBS model (see appendix B), we obtain that τ can be written as a function of t , θ and λ , where θ is an angle that relates to the curvature of the SSBS arc in the interval (t, λ) , that is, the attenuation we want to impose to data with in-between amplitudes. Since we have $0 < \theta < \arccos\left((\lambda - t)/\sqrt{4\lambda^2 + (\lambda - t)^2}\right)$, the larger θ , the stronger the attenuation of the coefficients with amplitudes in (t, λ) . Hereafter, parameter θ will be called the *attenuation degree* and the SSBS functions are written, equivalently, in the form $\delta_{t,\theta,\lambda}$, where the bijection between (t, θ, λ) and (t, τ, λ) is detailed in appendix B (see in particular Eqs. (25) and (26) in the said appendix).

In practice, when t and λ are fixed, the foregoing makes it possible to control the attenuation degree we want to impose to the data in (t, λ) by choosing θ , a rather natural parameter. This interpretation of the SSBS parameters makes it easier to select convenient values of these parameters for practical applications. Summarizing, the estimation procedure is performed in three steps:

1. Fix the asymptotic attenuation t , the threshold λ and the attenuation degree θ of the SSBS function.
2. Compute the corresponding value of τ from Eq. (26).
3. Shrink the data according to the SSBS function $\delta_{t,\tau,\lambda}$ of Eq. (2).

Some SSBS graphs are plotted in figure 4 for different values of the attenuation degree θ (threshold λ is fixed and the asymptotic attenuation parameter t is 0).

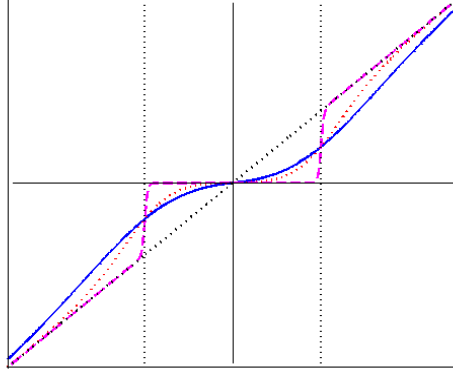


Figure 4: Shapes of SSBS functions $\delta_{\theta,\lambda}$ for different values of the attenuation degree θ : $\theta = \pi/6$ for the continuous (blue) curve, $\theta = \pi/4$ for the dotted (red) curve, and $\theta = \pi/3$ for the dashed (magenta) curve.

3 Penalty functions associated to SSBS in a regularization problem

Consider signal estimation by using the penalized least squares approach given in [12]. In this reference, the signal estimation is addressed by considering a penalty function $q_\lambda = q_\lambda(\cdot)$ and by looking for the vector \mathbf{d} that minimizes

$$\|\mathbf{d} - \mathbf{c}\|_{\ell_2}^2 + 2 \sum_{i=1}^N q_\lambda(|d_i|). \quad (3)$$

In [26], the unification between shrinkages and regularization procedures is discussed. It follows from this reference that shrinkages and regularization procedures are linked in the sense that a shrinkage function corresponds to a regularization problem with a specific penalty function. This correspondence is made more precise by the following proposition:

Proposition 1

[26, Proposition 3.2].

Let δ be a real valued thresholding function that is increasing, antisymmetric, such that $0 \leq \delta(x) \leq x$ for $x \geq 0$ and $\delta(x)$ tends to infinity as x tends to infinity. Then, there exists a continuous positive penalty function q , with $q(|x|) \leq q(|y|)$ whenever $|x| \leq |y|$, such that $\delta(z)$ is the unique solution of the minimization problem $\min_t (t - z)^2 + 2q(|t|)$ for every z at which δ is continuous. The penalty q associated with δ is given by

$$q(x) = \int_0^x (r(z) - z) dz, \quad (4)$$

for $x \geq 0$, where r is the generalized inverse of δ : $r(x) = \sup\{z \mid \delta(z) \leq x\}$.

An SSBS function $\delta_{t,\tau,\lambda}$ satisfies the assumptions of proposition 1. It follows that the shrinkage obtained by using a function $\delta_{t,\tau,\lambda}$ can be seen as a regularization approximation with a continuous positive penalty function. The following characterizes the penalty function associated to $\delta_{\tau,\lambda} = \delta_{0,\tau,\lambda}$.

Proposition 2

The shrinkage obtained by using an SSBS function $\delta_{\tau,\lambda}$ can be seen as a regularization approximation with penalty function $q_{\tau,\lambda}$, where $q_{\tau,\lambda}$ is the function defined for every $x \geq 0$ by

$$q_{\tau,\lambda}(x) = \frac{1}{\tau} \int_0^x \mathcal{L}(\tau z e^{-\tau(z-\lambda)}) dz, \quad (5)$$

with \mathcal{L} being the Lambert function defined as the inverse of the function: $t \geq 0 \longrightarrow t e^t$.

PROOF

Since SSBS functions are continuous and strictly increasing functions, the generalized inverse of any SSBS function $\delta_{\tau,\lambda}$ is the inverse, denoted $r_{\tau,\lambda}$, of this SSBS function. From proposition 1, the penalty associated with $\delta_{\tau,\lambda}$ is then

$$q_{\tau,\lambda}(x) = \int_0^x (r_{\tau,\lambda}(z) - z) dz. \quad (6)$$

Now, because the SSBS function $\delta_{\tau,\lambda}$ is continuous, strictly increasing and antisymmetric, its inverse $r_{\tau,\lambda}$ has the form

$$r_{\tau,\lambda}(z) = zG(z), \quad (7)$$

for every real value z and where G is such that

$$G(z) = 1 + e^{-\tau(|z|G(z)-\lambda)}.$$

Therefore, $G(z) > 1$ for any real value z . We thus have

$$(G(z) - 1) e^{\tau(|z|(G(z)-1))} = e^{-\tau(|z|-\lambda)},$$

which is also equivalent to

$$\tau|z|(G(z) - 1) e^{\tau(|z|(G(z)-1))} = \tau|z|e^{-\tau(|z|-\lambda)}.$$

It follows that

$$\tau|z|(G(z) - 1) = \mathcal{L}\left(\tau|z|e^{-\tau(|z|-\lambda)}\right),$$

which leads to

$$G(z) = 1 + \mathcal{L}\left(\tau|z|e^{-\tau(|z|-\lambda)}\right) / (\tau|z|) \quad (12)$$

for $z \neq 0$. Taking into account (7), (12) and the fact that $r_{\tau,\lambda}(0) = 0$ since $\delta_{\tau,\lambda}(0) = 0$, we obtain

$$r_{\tau,\lambda}(z) = z + \text{sgn}(z) \mathcal{L}\left(\tau|z|e^{-\tau(|z|-\lambda)}\right) / \tau \quad (13)$$

for any real value z . The result then follows by injecting Eq. (13) into Eq. (6). \square

From proposition 2, we derive that for every real value x , the value $\delta_{\tau,\lambda}(x)$ is the unique solution of the minimization problem $\min_d \left((x-d)^2 + 2q_{\tau,\lambda}(|d|) \right)$ where $q_{\tau,\lambda}$ is given by Eq. (5).

The shape of the SSBS penalty $q_{\tau,\lambda}(|x|)$ is given for fixed λ and several values of τ in figure 5; the penalties displayed in this figure are those associated with the SSBS functions of figure 4.

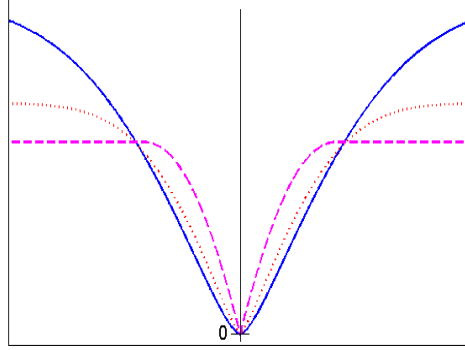


Figure 5: Penalty functions associated with the SSBS functions of figure 4 in the regularization problem of Eq. (3). The attenuation degree θ is: $\theta = \pi/6$ for the continuous (blue) curve, $\theta = \pi/4$ for the dotted (red) curve, and $\theta = \pi/3$ for the dashed (magenta) curve.

It follows that the penalty associated with an SSBS function is regular everywhere. This regularity depends on the SSBS shape. When the attenuation degree is small, the variability of treatment among data is reduced: the shape of the penalty function is more regular. In contrast, a large attenuation degree amplifies the variability of treatment: the slope of the penalty shape is strong for small data and tends to be quasi-null for large data (small data are strongly shrunk whereas large data are approximately kept). Note, by comparing figure 5 with [26, Figure 3], that the larger the SSBS attenuation degree, the closer to the hard penalty the SSBS penalty is.

4 Unification of basic thresholds

In [4], it is shown that soft thresholding estimation of signals in the wavelet domain can be improved by using the detection thresholds. In this section, we derive some properties of the detection thresholds proposed in [4]. In particular, Section 4.1 highlights that standard minimax and universal thresholds correspond to detection thresholds associated with different sparsity degrees and Section 4.2 provides some detection thresholds suitable for selecting the significant wavelet coefficients.

4.1 Detection thresholds

Consider the following decision problem with binary hypothesis model $(\mathcal{H}_0, \mathcal{H}_1)$, where $\mathcal{H}_0 : c_i \sim \mathcal{N}(0, \sigma^2)$ versus $\mathcal{H}_1 : c_i = d_i + \epsilon_i, |d_i| \geq \alpha > 0, \epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Assume that the *a priori* probability of occurrence of hypothesis \mathcal{H}_1 is less than or equal to some value $p \leq 1/2$. Then, for deciding \mathcal{H}_0 versus \mathcal{H}_1 , the thresholding test with threshold height

$$\lambda_D(\alpha, p) = \sigma \xi(\alpha/\sigma, p), \quad (14)$$

where

$$\xi(\alpha, p) = \frac{\alpha}{2} + \frac{1}{\alpha} \left[\ln \frac{1-p}{p} + \ln \left(1 + \sqrt{1 - \frac{p^2}{(1-p)^2} e^{-\alpha^2}} \right) \right], \quad (15)$$

has the same sharp upper bound for its probability of error than the Bayes test with the least probability of error (see [4] for details).

Parameter p reflects the presence (quantity) of significant coefficients of the signal amongst the noisy coefficients. Assuming that p is less than or equal to $1/2$ ensures that the representation of the signal is, at least, sparse in the weak sense. Parameter α can be seen as the minimum amplitude considered to be significant for a signal coefficient. Parameters p and α can thus be used to measure the sparsity degree of the signal representation (see [4], [15]).

The following proposition makes it possible to unify the minimax, universal, and detection thresholds.

Proposition 3

For any positive real value $\eta \geq \sigma$, there exist $\alpha_0 > 0$ and p_0 , with $0 \leq p_0 \leq 1/2$, such that

$$\lambda_D(\alpha_0, p_0) = \eta. \quad (16)$$

PROOF

The result simply follows by noting that the function ξ is continuous, positive, $\lim_{\alpha \rightarrow 0} \xi(\alpha, 1/2) = 1$, and $\lim_{\alpha \rightarrow +\infty} \xi(\alpha, p) = +\infty$ for any p such that $0 < p \leq 1/2$. \square

Let $\lambda_u(N) = \sigma \sqrt{2 \ln N}$, the so-called universal threshold. This threshold reflects the maximum amplitude of the noise coefficients. Indeed, since $\epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ for $i = 1, 2, \dots, N$, then it follows from [27, p. 187] or [28, p. 454], that

$$\lim_{N \rightarrow +\infty} \mathbb{P} \left[\lambda_u(N) - \frac{\sigma \ln(\ln N)}{\ln N} \leq \max \{ |\epsilon_i|, 1 \leq i \leq N \} \leq \lambda_u(N) \right] = 1. \quad (17)$$

Thus, the maximum amplitude of $\{\epsilon_i\}_{1 \leq i \leq N}$ has a strong probability of being close to the universal threshold when N is large.

From proposition 3, it follows that for any $N > 2$, there exist some values α, p such that $\lambda_D(\alpha, p) = \lambda_u(N)$. Figure 6 shows the level curves $\lambda_D(\alpha, p) = \lambda_u(N)$ for different values of N . It appears that large values of α are associated with small values of p (strong sparsity) and *vice versa* (weak sparsity).

The same remark (as for the universal threshold) holds true for the *minimax threshold*. The minimax threshold $\lambda_m(N)$ is defined as the largest value λ among the values attaining a minimax risk bound given in [1].

Assume that the noise standard deviation σ is fixed. Let $\sigma = 1$ for the sake of simplicity and we consider that noise is AWGN in the time/spatial domain. For a given signal, it is well-known

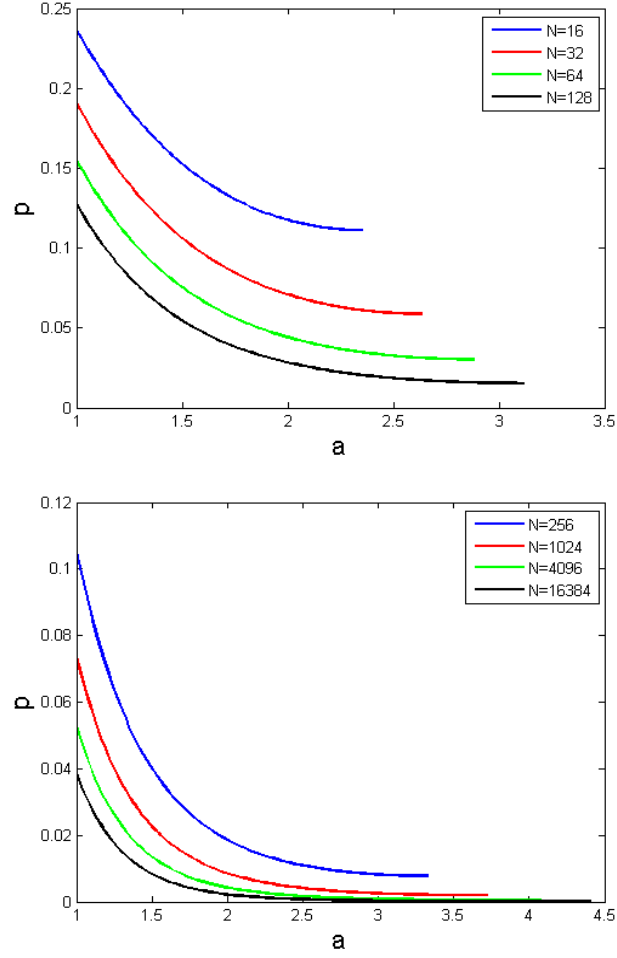


Figure 6: Level curves $\lambda_D(a, p) = \sqrt{2 \ln N}$ for different values of N .

that varying the threshold height λ used for selecting the signal wavelet coefficients yields different regularities for the signal estimate, larger threshold heights resulting in smoother estimates [1], [3]. Now, one can look for the class of signals (in terms of their sparsity measures a and p) that correspond to the same threshold height λ . We have from Eqs. (15) and (14) that

$$\lambda_D(a, p) = ap \left(\frac{1}{2p} + \vartheta(a, p) \right)$$

where

$$\vartheta(\alpha, p) = \frac{1}{\alpha^2 p} \left[\ln \frac{1-p}{p} + \ln \left(1 + \sqrt{1 - \frac{p^2}{(1-p)^2} e^{-\alpha^2}} \right) \right].$$

Since $0 < p \leq 1/2$, it follows that $\lambda_D(\alpha, p) \geq \alpha p$. Thus, for any $\eta \geq 1$, we have

$$\lambda_D(\alpha, p) = \eta \implies \alpha p \leq \eta \quad (18)$$

Eq. (18) confirms the fact that for a given value η (threshold height), the “minimum significant amplitude for the signal” α and the “proportion of significant signal coefficients” p characterising the surface $\lambda_D(\alpha, p) = \eta$ cannot be both arbitrarily large because of the constraint $\alpha p \leq \eta$.

In addition, we have $\Gamma(\alpha, p) \geq \frac{1}{\alpha^2 p} \ln \frac{1-p}{p}$ and thus $\lambda_D(\alpha, p) \geq \frac{1}{\alpha} \ln \frac{1-p}{p}$. By setting

$$p^* = 1 / \ln \frac{1-p}{p},$$

we have

$$\lambda_D(\alpha, p) = \eta \implies \alpha p^* \geq \frac{1}{\eta} \quad (19)$$

Note that p^* is a positive and strictly increasing function of p ($0 < p \leq 1/2$) and p^* tend to 0 when p tend to 0. Thus, from Eq. (19), it follows that for any given threshold height $\eta \geq 1$, the “minimum significant amplitude for the signal” α and the “proportion of significant signal coefficients” p characterising the surface $\lambda_D(\alpha, p) = \eta$ cannot be both arbitrarily small because of the constraint $\alpha p^* \geq 1/\eta$.

Summarizing, the class of signals (identified by the sparsity measures α and p) that admit the same threshold height is such that the uncertainty relations given by Eqs. (18) and (19) are satisfied. This class does not contain non-sparse signals since both α and p cannot be arbitrary large or small (as a matter of justification, note that thresholding is not very appropriate for estimating non-sparse signals). This class contains sparse signals with different sparsity degrees: when α is large and p is small, we will say that the signal under consideration admit a strongly

sparse representation; in contrast, a weak sparse signal representation is such that α is small and p is large (with the upper-bound for p fixed to be $1/2$: $p \leq 1/2$).

The above uncertainty relations thus allows for classifying signals according to their sparsity measures α and p and it follows that strongly and weakly sparse signals have the same threshold height η whenever their sparsity measures α and p are such that $\lambda_D(\alpha, p) = \eta$ (see figure 6 for some examples of level curves such that $\lambda_D(\alpha, p) = \eta$).

4.2 Detection thresholds adapted to the wavelet decomposition

Detection thresholds are well-adapted to estimate wavelet coefficients corrupted by AWGN because of the sparsity of the wavelet transform [4]. Moreover, these thresholds are adaptable to the wavelet transform decomposition schemes: sparsity ensures that for reasonable resolution levels, signal coefficients are less present than noise coefficients among the detail wavelet coefficients and that signal coefficients have large amplitudes (in comparison to noise coefficients).

More precisely, it is known that for smooth or piecewise regular signals, the proportion of significant coefficients, which plays a role similar to that of p , increases with the resolution level [28, Section 10.2.4, p. 460]. Therefore, if we can give, first, upper-bounds $(p_j)_{j=1,2,\dots,J}; p_j \leq 1/2$ for every decomposition levels $j = 1, 2, \dots, J$, and second, lower-bounds $(\alpha_j)_{j=1,2,\dots,J}$ for the amplitudes of the significant wavelet coefficients, then we can derive level-dependent detection thresholds that can select significant wavelet coefficients at every resolution level. Since significant information tends to be absent among the first resolution level detail wavelet coefficients, it is reasonable to set $\alpha_1 = \sigma \sqrt{2 \ln N}$, that is the universal threshold. Now, when the resolution level increases, it follows from [28, Theorem 6.4] that a convenient choice for $\alpha_j, j > 1$ is $\alpha_j = \alpha_1 / \sqrt{2^{j-1}}$ when the signal of interest is smooth or piecewise regular.

In addition, since noise tends to be less present when the resolution level increases, p_j must be an increasing function of j . Note that detection thresholds are defined for $p_j \leq 1/2$. It is

thus necessary to stop the shrinkage at a resolution level J for which p_J is less than or equal to $1/2$. We propose the use of exponentially or geometrically increasing sequences for the values $(p_j)_{j=1,2,\dots,J}$ since p_1 must be a very small value (significant information tends to be absent among the first resolution level detail wavelet coefficients) and the presence of significant information increases significantly as the resolution level increases. In the following, we consider a sequence $(p_j)_{j=1,2,\dots,J}$ such that $p_{j+1} = (p_j)^{1/\mu}$ with $\mu > 1$.

Summarizing, we consider the thresholds $\lambda_D(a_j, p_j)$, where λ_D is defined by Eq. (14) and (a_j, p_j) , for $j = 1, 2, \dots, J$ are given by

$$a_j = \sigma \sqrt{\ln N} / 2^{j/2-1}, \quad (20)$$

and

$$p_j = 1/2^{\mu^{J-j}}. \quad (21)$$

5 SSBS and detection thresholds in practice

This section provides experimental results which highlight that SSBS allows for noise reduction with preservation of structural details of images. We first discuss the calibration of the SSBS parameters and assess the SSBS performance when noise is AWGN (Section 5.1). We then apply SSBS to the denoising of SAR (Synthetic Aperture Radar) images (Section 5.2).

Experimental tests are carried out by using the Stationary Wavelet Transform (SWT) [29]. The maximum decomposition level is fixed to $J = 4$. The SWT has appreciable properties in denoising. Its redundancy makes it possible to reduce residual noise and some possible artifacts incurred by the translation sensitivity of the orthonormal wavelet transform.

5.1 Denoising images corrupted by synthetic AWGN

In this section, we consider a class of test images corrupted by synthetic AWGN. We use the PSNR and the SSIM index in order to assess the quality of a denoised image. The PSNR (Peak Signal-to-Noise Ratio, in deciBel unit, dB) refers to the Mean Square Error (MSE) and is given by

$$\text{PSNR} = 10 \log_{10} (255^2 / \text{MSE}), \quad (22)$$

The SSIM (Structural SIMilarity) index [30] is a perceptual measure that compares patterns of pixel intensities for images, on the basis of the local luminance and contrast of the analyzed pixels. Let x and y be two data vectors assumed to contain non-negative values only and representing the pixel values to be compared. The luminance and the contrast of these pixels are estimated by the mean and the standard deviation of x and y , respectively. The SSIM index between x and y is then given by [30]:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (23)$$

where μ_x, σ_x , (resp. μ_y, σ_y) are the mean and standard deviation of x (resp. y) and σ_{xy} designate the covariance between x and y . The local statistics $\mu_x, \mu_y, \sigma_x, \sigma_y$ and σ_{xy} are computed within a window with size 11×11 and the pixel values in this window are normalized by using a unit sum circular-symmetric Gaussian weighting function (see [30] for details). We also use the constants C_1 and C_2 suggested by the authors in [30]: $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$ with $K_1 = 0.01$ and $K_2 = 0.03$, where L is the dynamic range of the pixel values ($L = 255$ for 8-bit grayscale images). The SSIM index of two images is then the average value of the different SSIM indices obtained by sliding the local window over the entire image.

5.1.1 Preliminary tests

We first run some preliminary tests in order to choose the SSBS parameters and analyze the sensitivity of these parameters. We consider the standard ‘House’, ‘Barbara’ and ‘Lena’ images,

which are very popular in the image denoising community. The tests are based on an analysis of the means and variances of the PSNRs and SSIMs computed over 10 noise realizations, when the test images are denoised by the SSBS method. The tested images are corrupted by AWGN with standard deviation $\sigma = 5, 10, 15, 20, 25, 30, 35$ and the SWT is computed by using different Daubechies, spline, and symlet wavelet filters.

Given a fixed global threshold that can be either the universal threshold (λ_u), the minimax threshold (λ_m) or the universal-detection threshold (λ_{ud} , obtained by setting $\alpha = \sigma\sqrt{2\ln N}$ and $p = 1/2$ in Eq. (14), see [4] for the properties of this threshold), the tests suggest using small (resp. large) asymptotic attenuation t and attenuation degree θ when the noise level is small (resp. large). The same remark holds for the level-dependent detection thresholds $\lambda_D(j) = \lambda_D(\alpha_j, p_j)$ defined by Eq. (14), where (α_j, p_j) , for $j = 1, 2, \dots, J$ are given by Eqs. (20) and (21). The value $\mu = 2.35$ in Eq. (21) tends to be a good compromise for the different test images used, when we assume that $p_j = 1/2$.

Among the global thresholds, the minimax and the universal-detection thresholds outperform the universal threshold, and the universal-detection threshold tends to be more performant than the minimax threshold, especially when the noise level is not very large. The level-dependent thresholds perform better than the global thresholds described above when the noise standard deviation is larger than 10. In addition, the preliminary tests show that reasonable asymptotic attenuation and attenuation degree parameters for SSBS are

- $0 \leq t \leq \sigma/10$ and $0 < \theta \leq \pi/10$ when the noise standard deviation σ is less than 10,
- $0 \leq t \leq \sigma/5$ and $\pi/10 \leq \theta \leq \pi/6$ when $10 \leq \sigma < 20$,
- $\sigma/10 \leq t \leq \sigma/3$ and $\pi/6 \leq \theta \leq \pi/4$ when σ is larger than or equal to 20.

For fixed parameters, we also test the sensitivity of the SSBS method according to the wavelet filters used. It follows that, for a given wavelet family, there is no significant variability of the

results with respect to the length of the wavelet filter used. In addition, there is no significant difference between the results obtained whatever the wavelet family used, provided that the length of the filters remain approximately of the same order.

5.1.2 SSBS denoising performance

In this section, we compare SSBS and BLS-GSM denoising performance. The BLS-GSM of [8] (free MatLab software ¹) is a parametric method using redundant wavelet transform and models neighbourhoods of wavelet coefficients with Gaussian vectors multiplied by random positive scalars. BLS-GSM also takes into account the orientation and the interscale dependencies of the wavelet coefficients. It is actually the best parametric method using redundant wavelet transform.

We present in tables 1, 2 and 3, some experimental results (PSNRs and SSIMs performed by the SSBS) when the tested image is the ‘Boat’ image (the variances are very small in comparison with the mean PSNRs). The results are similar with other standard images such as the ‘Peppers’ and the ‘Fingerprint’ images. We focus on the case where the noise standard deviation σ is between 5 and 15. Indeed, this case is of interest since when $\sigma \leq 3$, noise is often non-perceptible and noise observed in images processed by modern acquisition systems is often moderate. We run the SSBS denoising procedure described above. The SWT is computed by using the biorthogonal spline wavelet with order 3 for the decomposition and order 1 for the reconstruction (‘bior1.3’ in Matlab Wavelet toolbox). Several values for the SSBS parameters are tested. These values are chosen according to the recommendations made in the previous section. We also present in the captions of tables 1, 2 and 3, the PSNRs and SSIMs achieved by the BLS-GSM method of [8].

According to the experimental results presented in tables 1, 2 and 3, the performance of

¹available at <http://decsai.ugr.es/~javier/denoise/software/index.htm>

Table 1: Means of the PSNRs and SSIMs computed over 10 noise realizations, when denoising the ‘Boat’ image by the SSBS method. The SSBS parameters (t, θ, λ) used are given in the table. The PSNR and the SSIM performed by the BLS-GSM method are 36.72 dB and 0.929 respectively.

$\sigma = 5$									
PSNR									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	36.16	36.53	36.28	36.01	36.48	36.14	35.84	36.38	35.99
$\theta = \pi/12$	36.23	36.55	36.32	36.09	36.52	36.19	35.92	36.44	36.04
$\theta = \pi/10$	36.25	36.54	36.30	36.12	36.52	36.16	35.96	36.45	36.03
$\theta = \pi/8$	36.18	36.48	36.20	36.05	36.47	36.06	35.92	36.40	35.93
$\theta = \pi/6$	35.94	36.30	35.87	35.82	36.28	35.75	35.68	36.24	35.64
SSIM									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	0.928	0.933	0.930	0.926	0.932	0.928	0.923	0.931	0.926
$\theta = \pi/12$	0.928	0.933	0.930	0.926	0.933	0.928	0.924	0.931	0.925
$\theta = \pi/10$	0.928	0.934	0.929	0.926	0.933	0.926	0.923	0.931	0.924
$\theta = \pi/8$	0.926	0.933	0.926	0.924	0.932	0.923	0.921	0.931	0.921
$\theta = \pi/6$	0.921	0.930	0.919	0.918	0.929	0.916	0.915	0.928	0.913

Table 2: Means of the PSNRs and SSIMs computed over 10 noise realizations, when denoising the ‘Boat’ image by the SSBS method. The SSBS parameters (t, θ, λ) used are given in the table. The PSNR and the SSIM performed by the BLS-GSM method are 33.48 dB and 0.878 respectively.

$\sigma = 10$									
PSNR									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	32.25	32.38	32.55	32.20	32.45	32.52	32.10	32.47	32.46
$\theta = \pi/12$	32.41	32.46	32.70	32.36	32.55	32.65	32.28	32.59	32.59
$\theta = \pi/10$	32.58	32.55	32.84	32.51	32.65	32.78	32.42	32.69	32.70
$\theta = \pi/8$	32.70	32.62	32.92	32.61	32.72	32.85	32.51	32.77	32.76
$\theta = \pi/6$	32.66	32.63	32.84	32.55	32.72	32.76	32.44	32.76	32.64
SSIM									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	0.851	0.849	0.860	0.853	0.853	0.861	0.853	0.856	0.861
$\theta = \pi/12$	0.856	0.853	0.864	0.857	0.857	0.865	0.856	0.859	0.865
$\theta = \pi/10$	0.861	0.857	0.869	0.861	0.861	0.869	0.860	0.863	0.868
$\theta = \pi/8$	0.865	0.862	0.872	0.864	0.865	0.871	0.861	0.867	0.869
$\theta = \pi/6$	0.864	0.865	0.870	0.861	0.867	0.867	0.857	0.868	0.864

Table 3: Means of the PSNRs and SSIMs computed over 10 noise realizations, when denoising the ‘Boat’ image by the SSBS method. The SSBS parameters (t, θ, λ) used are given in the table. The PSNR and the SSIM performed by the BLS-GSM method are 31.63 dB and 0.839 respectively.

$\sigma = 15$									
PSNR									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	29.84	29.78	30.21	29.87	29.95	30.28	29.86	30.06	30.29
$\theta = \pi/12$	30.08	29.94	30.43	30.09	30.11	30.49	30.08	30.24	30.48
$\theta = \pi/10$	30.35	30.11	30.68	30.34	30.30	30.70	30.30	30.42	30.69
$\theta = \pi/8$	30.60	30.31	30.91	30.57	30.48	30.89	30.50	30.60	30.85
$\theta = \pi/6$	30.74	30.47	31.01	30.66	30.64	30.95	30.54	30.73	30.86
SSIM									
	$t = 0$			$t = \sigma/10$			$t = \sigma/5$		
	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
$\theta = \pi/14$	0.770	0.758	0.783	0.776	0.768	0.790	0.781	0.777	0.794
$\theta = \pi/12$	0.781	0.766	0.794	0.786	0.776	0.800	0.790	0.785	0.803
$\theta = \pi/10$	0.794	0.777	0.807	0.797	0.786	0.811	0.800	0.794	0.814
$\theta = \pi/8$	0.808	0.789	0.821	0.809	0.798	0.822	0.809	0.805	0.822
$\theta = \pi/6$	0.817	0.802	0.828	0.815	0.810	0.827	0.811	0.814	0.825

SSBS and BLS-GSM are of the same order, both in terms of PSNR and SSIM. The BLS-GSM yields a PSNR slightly higher than the SSBS, the difference in PSNR between SSBS and BLS-GSM being less than 1 dB. The (best) SSBS yields higher SSIM quality index when the noise standard deviation is 5 and the BLS-GSM SSIM is slightly higher when the noise standard deviation is 10, 15.

From these results, it follows that SSBS and BLS-GSM are comparable, both in terms of PSNR and SSIM quality index. In comparison with BLS-GSM, the advantage of SSBS is then its extreme algorithmic simplicity. Indeed, SSBS can be seen as a weighting function that simply applies to the wavelet coefficients, whereas BLS-GSM is computationally expensive and cannot be used in an operational context (see [13] for an appreciation of the BLS-GSM computing time). Note that, in contrast with the BLS-GSM and other denoising methods such as the SURELET of [13], which yields performances of the same order as BLS-GSM, SSBS uses neither interscale nor intrascale predictors. These predictors can be included in the shrinkage process for SSBS and they can certainly allow for better denoising results. However, such predictors are generally specific to the wavelet transform used and, in this respect, they are detrimental to the portability of a method. For this reason, we do not consider interscale and intrascale predictors in this paper. Note also that SSBS might be adapted to other *a priori* knowledge by using some directional processing similar to that employed by BLS-GSM. However, when the noise level is small (which is the case of interest in practical applications such as SAR denoising) then it follows, from the SSIM indices given in table 1, that SSBS, without any additional *a priori* information, guarantees better preservation of structural information contained in the images than BLS-GSM. In other words, SSBS guarantees noise reduction without impacting significantly the signal characteristics (figures 7, 8 and 9 provide SSBS and BLS-GSM denoisings for the ‘Boat’ image as an illustration). In addition, it follows from tables 1, 2 and 3 that SSBS denoising performance is not really affected by slightly different SSBS parameter values. This highlights the robustness of the method

Noisy image
AWGN, $\sigma = 5$
PSNR=34.1514 dB
SSIM=0.8853



Noisy image
AWGN, $\sigma = 15$
PSNR=24.6090 dB
SSIM=0.5364



Figure 7: Noisy 'Boat' image corrupted by AWGN with standard deviation σ .

Denoising
SSBS $\delta_0, \frac{\pi}{10}, \lambda_{ud}$
PSNR=36.5719 dB
SSIM=0.9339



Denoising
BLS-GSM
PSNR=36.7129 dB
SSIM=0.9290



Figure 8: SSBS and BLS-GSM denoising of noisy 'Boat' image corrupted by AWGN with $\sigma = 5$.

Denoising
SSBS $\delta_{0, \frac{\pi}{6}, \lambda_D(j)}$
PSNR=30.9954 dB
SSIM=0.8274



Denoising
BLS-GSM
PSNR=31.6396 dB
SSIM=0.8394



Figure 9: SSBS and BLS-GSM denoising of noisy 'Boat' image corrupted by AWGN with $\sigma = 15$.

5.2 Denoising SAR images

In image processing, denoising is of interest, specifically for high resolution images such as biomedical ultrasonic or SAR images, for instance. In such images, the signal reflectance z is corrupted by speckle noise ϵ . Speckle noise is modeled as a correlated stationary process, multiplicative with the signal reflectance. The observation is then ϵz . We can write $\epsilon z = z + z(\epsilon - 1)$ so as to consider that a SAR image is the sum of the signal reflectance z and a signal-dependent noise $z(\epsilon - 1)$. In this signal-dependent noise case, the performance of a wavelet shrinkage is not guaranteed to be as performant as in the AWGN case. Furthermore, we cannot guarantee that the signal-dependent noise wavelet coefficients can be rendered sufficiently iid Gaussian. However, for SAR images, speckle removal (despeckling) by wavelet shrinkage has been successfully addressed by several authors, mostly in the case of parametric models (see [31], [32], [33], among others). We thus carry out some experiments on SAR images to assess the relevance of SSBS in this case.

According to Section 5.1.2 above, the advantage of SSBS in the context of SAR image despeckling is that SSBS is a simple and performant non-parametric method that allows for different levels of noise reduction, the noise reduction being smoothly adjustable thanks to the flexibility of the SSBS parameters. Because of this flexibility, we can investigate noise reduction instead of full denoising for the SAR images. Indeed, because speckle contains much information, it must not be considered as pure noise. More precisely, we do not wish to fully remove speckle, but we want to reduce the variability due to it, without impacting the structural information of the SAR data.

- **[Quality criterion]** Since the reference (noise free) image is not available, we use, as a quality criterion, the Equivalent Number of Looks (ENL) for the SAR images combined

with the “method-noise” approach of [34]. The ENL of a SAR intensity image x is given by

$$\text{ENL}(x) = \frac{(\text{Mean}(x))^2}{\text{Variance}(x)}$$

In an homogeneous region, and for a speckle free image, this quantity will often be very large since homogeneity assumes small variability among the data. In contrast, in presence of large variability among the data, this quantity will often be small. When speckle is fully developed in an homogeneous region of a SAR image, then ENL is a good estimate of the number of looks used to form the SAR intensity image.

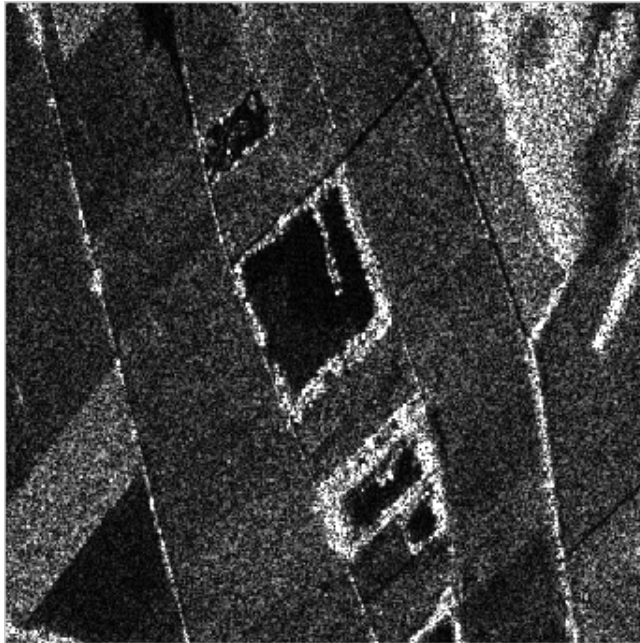
In addition, and since ENL does not make it possible to measure the signal distortion caused by the denoising, we also use another quality criterion: a variant of the ‘method-noise’. The method-noise [34] simply involves analyzing the difference between the original (noisy) image and the denoised image. According to this method, denoising quality is appreciated by checking the structural contents of the method-noise (difference) image. Assuming that noise is additive, this method-noise image looks like a pure noise image for methods capable of reducing noise without impacting the structural contents of the images. In the case of SAR despeckling, we consider the following variant of the method-noise. This variant is hereafter called the ‘method-noise ratio’ and involves computing the ratio between the noisy image and the denoised image. Indeed, the difference is less informative than the ratio because in the ideal case where the estimate equals the reflectance z , the difference consists of the signal-dependent noise $z(\epsilon - 1)$ whereas the ratio is exactly the speckle noise ϵ . For the above variant, speckle reduction can be considered to be more accurate, that is, to better preserve the structural contents of the signal reflectance, when the method-noise ratio image looks like pure speckle noise.

- **[Estimation of the noise standard deviation]** Let us denote by (j, α) where $j \in \{1, 2, \dots, J\}$ is the wavelet decomposition level and α denotes either the horizontal, the vertical or the

diagonal detail sub-image obtained from the SWT of a given input image. In order to estimate the noise standard deviation in a given sub-image (j, α) , we follow a strategy similar to that proposed in [35, Section 2.3]. Specifically, we use the standard MAD (Median of the Absolute Deviation) based estimator of the noise standard deviation [1]. The difference with [35, Section 2.3] is that we apply this estimator to a high-pass filtered version of the sub-image under consideration. More precisely, we decompose again this sub-image (j, α) by using a one-level DWT and apply the MAD based estimator on the diagonal detail DWT coefficients. The rationale is that outliers present in the sub-image (j, α) will be less present among the detail DWT coefficients of this sub-image, because of the high-pass filtering effect [28, Section 10.2.4, p. 459]. We use the diagonal detail coefficients because these coefficients usually contain less signal coefficients than vertical and horizontal detail coefficients. In particular, when the noise standard deviation is known, the tests we performed showed that its MAD based estimate computed on the basis of the diagonal detail coefficients tends to be more precise than the estimate obtained by averaging the 3 MAD based estimates computed from the horizontal, vertical and diagonal detail coefficients. We recall that the MAD based estimate of the noise standard deviation of a data set $x = (x_i)_i$ is given by $\bar{\sigma} = \text{Median}(|x|)/0.6745$. The robustness of the MAD is due to the fact that the median is not really affected by a small number of outliers and is not very sensitive to a small change in the data [1], [36].

Figure 10 provides the SAR images used for the experimental tests. Table 4 presents the ENLs obtained by SSBS denoising of these images (different SSBS parameters are used for the denoising). The ENLs are computed within a window located in an homogeneous region for the images under consideration (original images and denoised images). From table 4, it follows that the variability reduction in the SAR data is more effective for SSBS adjusted with the level-dependent detection thresholds than for SBBS adjusted with the minimax or the universal-

(a) ESAR Image



(b) ERS Image, © ESA

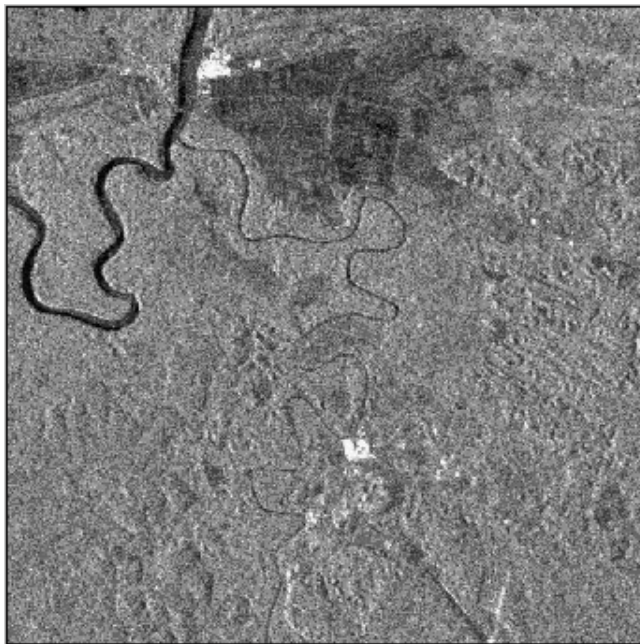


Figure 10: SAR images used the for experimental tests. (a): ESAR image with metric resolution; (b): ERS image with decametric resolution.

detection threshold.

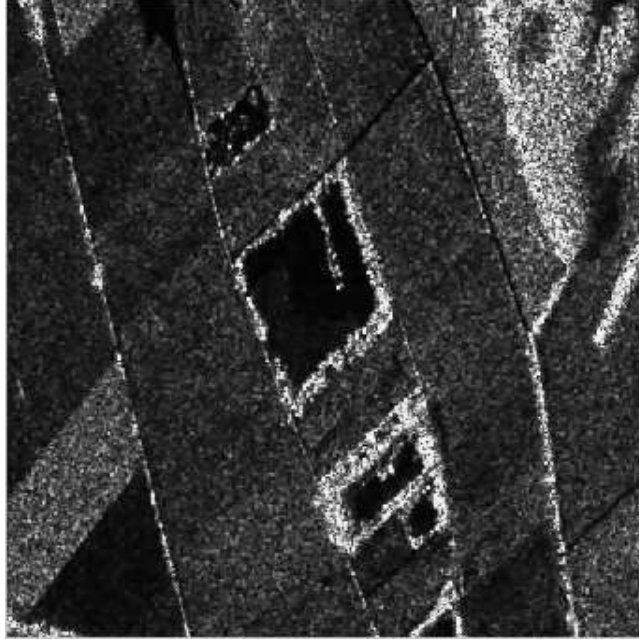
Examples of SAR SSBS denoisings are given in figures 11-(a) and 12-(a). The SSBS parameters used are indicated below the figure. The method-noise ratio images obtained from these denoisings are given in figures 13-(a) and 14-(a). As can be seen in these figures, the ratio between an original SAR image and its SSBS denoised version looks like speckle noise: it does not contain significant signal components. Thus, SSBS performs noise reduction without impacting significantly the signal characteristics: the non-informative speckle components are removed in homogenous areas whereas the speckle-like textural information contained in the original data is preserved. Figures 11-(b), 12-(b), 13-(b) and 14-(b) also provide experimental results for the BLS-GSM denoising of the SAR images of figure 10, for comparison with SSBS. The method-noise ratio images obtained from the BLS-GSM denoising still contain non-speckled signal components, as can be seen in figures 13-(b) and 14-(b): the rough denoising performed by BLS-GSM (see figures 11-(b) and 12-(b)) thus affects the structural information of the SAR

Table 4: ENLs obtained for SSBS denoising of the SAR images of figure 10. The SSBS parameter t is fixed to 0 for the denoising. The ENLs equal 3.0729 for the original ESAR image and 2.6664 for the original ERS image. The BLS-GSM denoising of these images yield ENLs equal 3.9285 for the ESAR image and 13.4375 for the ERS image.

$\theta = \pi/8$			$\theta = \pi/6$		
λ_m	λ_{ud}	$\lambda_D(j)$	λ_m	λ_{ud}	$\lambda_D(j)$
SSBS ESAR image denoising					
4.8164	4.0227	5.4051	4.6342	3.8766	5.3136
SSBS ERS image denoising					
4.6397	3.6993	4.8125	4.4943	3.5570	4.7546

data.

(a) ESAR Image
Denoising by using
SSBS $\delta_{0, \frac{\pi}{8}, \lambda_D(j)}$



(b) ESAR Image
Denoising by using
BLS-GSM

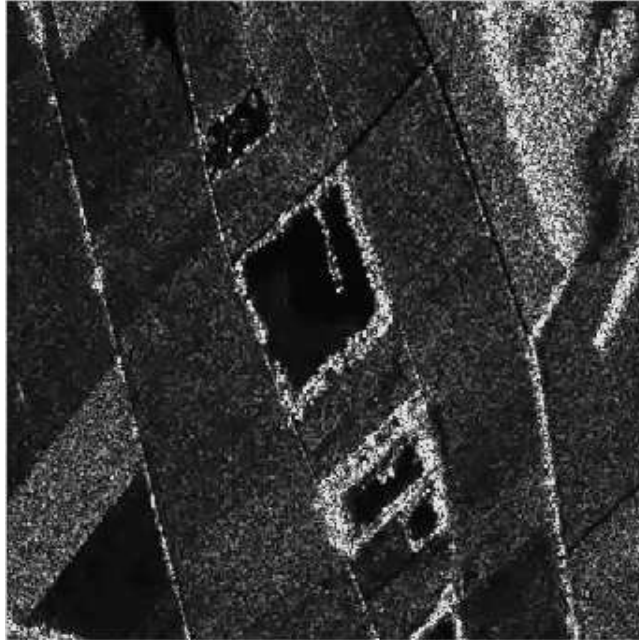
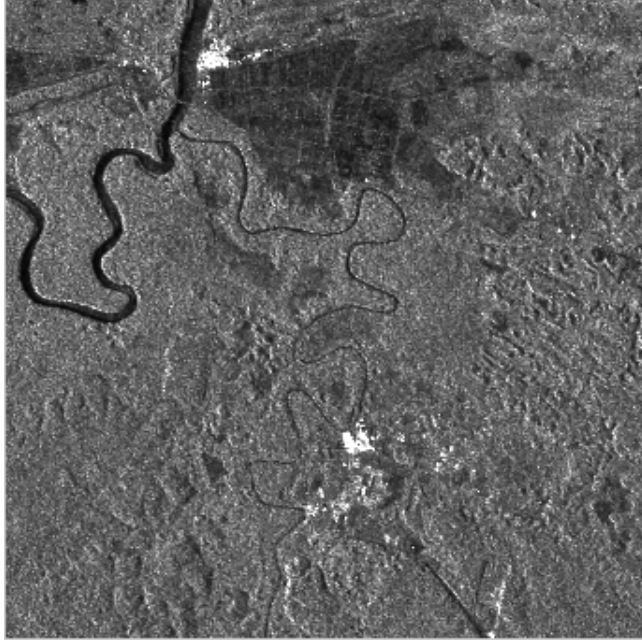


Figure 11: SSBS and BLS-GSM denoisings for the ESAR image of figure 10-(a).

(a) ERS Image, © ESA
Denoising by using
SSBS $\delta_{0, \frac{\pi}{8}, \lambda_D(j)}$



(b) ERS Image, © ESA
Denoising by using
BLS-GSM

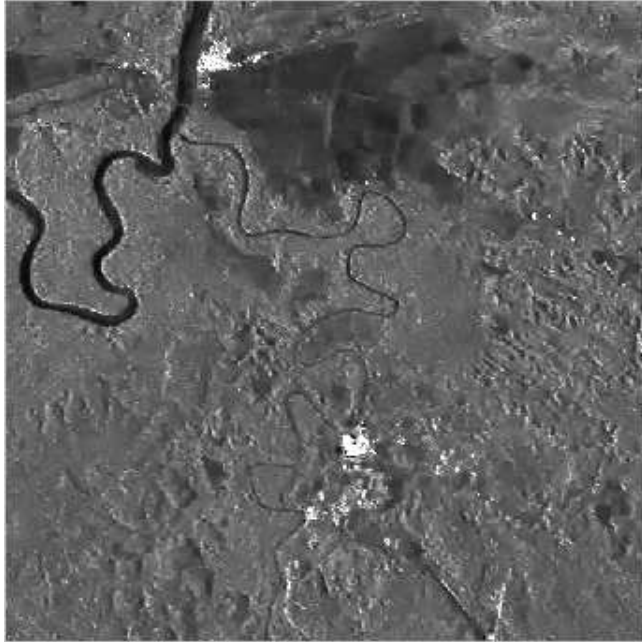
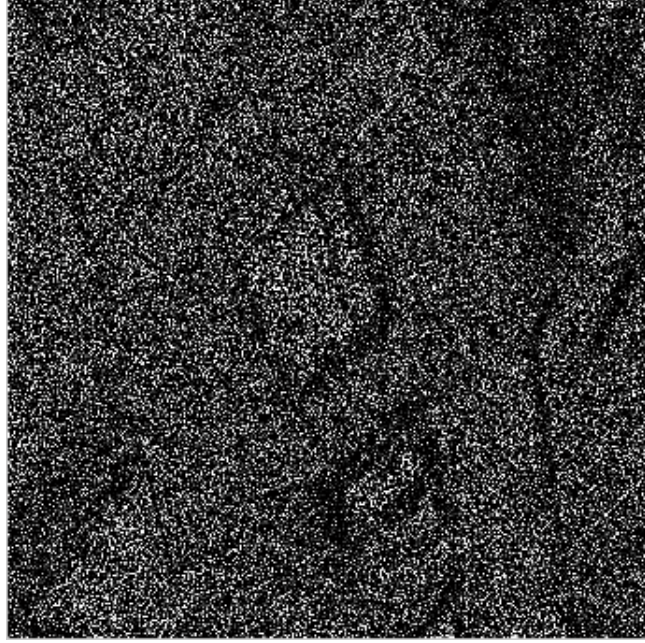


Figure 12: SSBS and BLS-GSM denoisings for the ERS image of figure 10-(b).

(a) ESAR Image
method-noise ratio with
SSBS $\delta_{0, \frac{\pi}{8}, \lambda_D(j)}$



(b) ESAR Image
method-noise ratio with
BLS-GSM

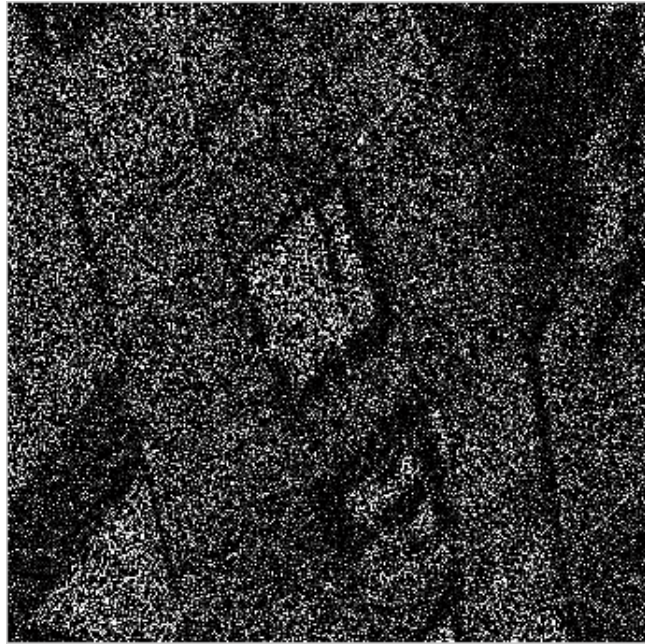


Figure 13: Image (a) is the ratio between the image of figure 10-(a) and the image of figure 11-(a). Image (b) is the ratio between the image of figure 10-(a) and the image of figure 11-(b).

(a) ERS Image, © ESA
method-noise ratio with
SSBS $\delta_{0, \frac{\pi}{8}, \lambda_D(j)}$



(b) ERS Image, © ESA
method-noise ratio with
BLS-GSM

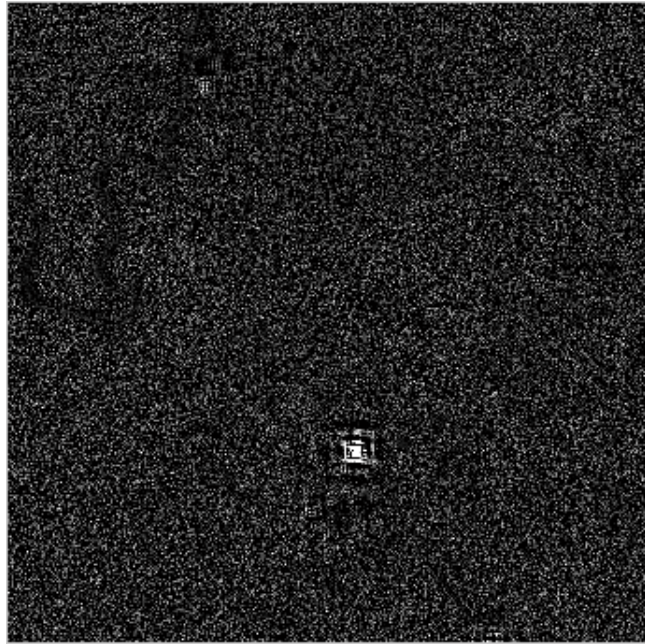


Figure 14: Image (a) is the ratio between the image of figure 10-(b) and the image of figure 12-(a). Image (b) is the ratio between the image of figure 10-(b) and the image of figure 12-(b).

To conclude this section, note that a sub-class of SSBS functions constitutes a class of invertible functions. These SSBS functions are those obtained by setting $t = 0$. For such an SSBS function, its inverse is given by Eq. (13). This allows for lossless denoising in the sense that one can retrieve an original image from its denoised version by simply applying the inverse denoising procedure: decompose the denoised image with the same wavelet transform as that initially used, apply the inverse of the SSBS function to the wavelet coefficients and reconstruct the original image by using the inverse wavelet transform. This lossless denoising might be relevant in many applications involving large databases. As a matter of fact, SAR, oceanography and medical ultrasonic sensors record many gigabits of data per day. These data (images) are mainly corrupted by speckle noise. Lossless despeckling of these databases is appealing since it is not essential to conserve a copy of the original database (thousands and thousands of gigabits recorded every year). By using SSBS denoising, one can thus retrieve an original image (when needed) by simply applying the inverse denoising procedure, which involves the inverse function of the SSBS used for the denoising.

6 Conclusion

Some noticeable properties of the SSBS functions and the detection thresholds have been highlighted. The SSBS functions are a family of smooth sigmoid based shrinkage functions that perform a penalized shrinkage. The standard hard and soft thresholding functions can be seen as degenerate SSBS functions. The properties of the SSBS functions have been addressed on the basis of the geometrical interpretation of their parameters. It follows that the SSBS functions are parameterized by 3 parameters that allow for controlling the attenuation degrees to be applied to small, median and large values, the shrinkage process being regular since non-degenerate SSBS functions are smooth functions.

On the other hand, this paper has also analyzed the properties of the detection thresholds. Detection thresholds depend on two parameters that describe the sparsity of the wavelet representation in terms of “minimum significant amplitude” for the signal and “probability of occurrence” of the significant signal coefficients in the sequence of the wavelet coefficients. It is shown that the universal and minimax thresholds are particular detection thresholds corresponding to different degrees of sparsity.

Finally, the use of detection thresholds for calibrating SSBS functions has been addressed. We have selected the SSBS detection thresholds on the basis of the known behavior of the wavelet coefficients for smooth and piecewise regular signals. The resulting shrinkage is performant for many images, including SAR images. The experimental results show that SSBS functions adjusted with these detection thresholds achieve denoising PSNRs and SSIMs comparable to those attained with the best parametric and computationally expensive method, the BLS-GSM of [8]. This performance is remarkable for a non-parametric method where no interscale or intrascale predictors are used to provide information about significant wavelet coefficients. The SSBS functions are thus suitable functions for noise reduction of large size signals and images. They also allow for a lossless denoising (due to the invertibility of a sub-class of SSBS functions), which can be relevant in many applications involving large databases.

An extension to this work could concern the estimation of the detection thresholds parameters α and p on the basis of the input noisy signal wavelet coefficients. This extension involves estimating the minimum significant wavelet coefficient for the signal and an upper bound on the probability of occurrence of significant wavelet coefficients. Combining this extension with predictors for detecting significant (small) wavelet coefficients could probably improve the SSBS performance.

Acknowledgement

The authors are very grateful to the reviewers for their insightful comments. They also would like to thank E. Pottier and J.P. Rudant for providing ESAR and ERS image data.

References

- [1] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425 – 455, Aug. 1994.
- [2] A. G. Bruce and H. Y. Gao, “Understanding waveshrink: Variance and bias estimation,” *Biometrika*, vol. 83, no. 4, pp. 727–745, 1996.
- [3] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200 – 1224, Dec. 1995.
- [4] A. M. Atto, D. Pastor, and G. Mercier, “Detection threshold for non-parametric estimation,” *Signal, Image and Video Processing, Springer*, vol. 2, no. 3, Sept. 2008.
- [5] E. P. Simoncelli and E. H. Adelson, “Noise removal via bayesian wavelet coring,” *IEEE Int. Conf. Image Proc. (ICIP)*, pp. 379–382, Oct. 1996.
- [6] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance,” *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146 – 158, Feb. 2002.
- [7] L. Şendur and I. V. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Transactions on Signal Processing*, vol. 11, pp. 2744–2756, Nov. 2002.

- [8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.
- [9] I. M. Johnstone and B. W. Silverman, "Empirical bayes selection of wavelet thresholds," *Annals of Statistics*, vol. 33, no. 4, pp. 1700–1752, 2005.
- [10] C. J. F. ter Braak, "Bayesian sigmoid shrinkage with improper variance priors and an application to wavelet denoising," *Computational Statistics and Data Analysis*, vol. 51, no. 2, pp. 1232 – 1242, 2006.
- [11] H. Y. Gao, "Waveshrink shrinkage denoising using the non-negative garrote," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, 1998.
- [12] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 939 – 955, Sep. 2001.
- [13] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, Mar. 2007.
- [14] A. M. Atto, D. Pastor, and G. Mercier, "Smooth sigmoid wavelet shrinkage for non-parametric estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Las Vegas, Nevada, USA, 30 march - 4 april, 2008.
- [15] D. Pastor and A. M. Atto, "Sparsity from binary hypothesis testing and application to non-parametric estimation," *European Signal Processing Conference, EUSIPCO*, Lausanne, Switzerland, August 25-29, 2008.

- [16] J. J. Benedetto and M. W. Frasier, *Wavelets : Mathematics and applications*. CRC Press, 1994, ch. 9 : Wavelets, probability, and statistics : Some bridges , by Christian Houdré, pp. 365–398.
- [17] J. Zhang and G. Walter, “A wavelet-based KL-like expansion for wide-sense stationary random processes,” *IEEE Transactions on Signal Processing*, vol. 42, no. 7, pp. 1737–1745, July 1994.
- [18] D. Leporini and J.-C. Pesquet, “High-order wavelet packets and cumulant field analysis,” *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 863–877, Apr. 1999.
- [19] A. M. Atto, D. Pastor, and A. Isar, “On the statistical decorrelation of the wavelet packet coefficients of a band-limited wide-sense stationary random process,” *Signal Processing*, vol. 87, no. 10, pp. 2320 – 2335, Oct. 2007.
- [20] A. M. Atto and D. Pastor, “Limit distributions for wavelet packet coefficients of band-limited stationary random processes,” *European Signal Processing Conference, EUSIPCO*, Lausanne, Switzerland, August 25-28, 2008.
- [21] P. Flandrin, “Wavelet analysis and synthesis of fractional brownian motion,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, Mar. 1992.
- [22] A. H. Tewfik and M. Kim, “Correlation structure of the discrete wavelet coefficients of fractional brownian motion,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 904–909, Mar. 1992.
- [23] R. W. Dijkerman and R. R. Mazumdar, “On the correlation structure of the wavelet coefficients of fractional brownian motion,” *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1609 – 1612, Sep. 1994.

- [24] T. Kato and E. Masry, "On the spectral density of the wavelet transform of fractional brownian motion," *Journal of Time Series Analysis*, vol. 20, no. 50, pp. 559–563, 1999.
- [25] P. F. Craigmile and D. B. Percival, "Asymptotic decorrelation of between-scale wavelet coefficients," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 1039 – 1048, Mar. 2005.
- [26] A. Antoniadis, "Wavelet methods in statistics: Some recent developments and their applications," *Statistics Surveys*, vol. 1, pp. 16 – 55, 2007.
- [27] S. M. Berman, *Sojourns and extremes of stochastic processes*. Wadsworth and Brooks/Cole, 1992.
- [28] S. Mallat, *A wavelet tour of signal processing, second edition*. Academic Press, 1999.
- [29] R. R. Coifman and D. L. Donoho, *Translation invariant de-noising*. Lecture Notes in Statistics, 1995, no. 103, pp. 125–150.
- [30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 – 612, April 2004.
- [31] J. R. Sveinsson and J. A. Benediktsson, "Speckle reduction and enhancement of sar images in the wavelet domain," *Geoscience and Remote Sensing Symposium, IGARSS*, vol. 1, pp. 63–66, May 1996.
- [32] H. Xie, L. E. Pierce, and F. T. Ulaby, "Sar speckle reduction using wavelet denoising and markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196 – 2212, Oct. 2002.

- [33] F. Argenti, T. Bianchi, and L. Alparone, “Multiresolution map despeckling of sar images based on locally adaptive generalized gaussian pdf modeling,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3385 – 3399, Nov. 2006.
- [34] A. Buades, B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490 – 530, 2005.
- [35] I. M. Johnstone and B. W. Silverman, “Wavelet threshold estimators for data with correlated noise,” *Journal of the Royal Statistical Society, Series B*, vol. 59, no. 2, pp. 319–351, 1997.
- [36] J. W. Müller, “Possible advantages of a robust evaluation of comparisons,” *Journal of Research of the National Institute of Standards and Technology*, vol. 105, no. 4, pp. 551–555, 2000.

A The soft and hard thresholding functions are degenerate SSBS functions

For fixed t and λ , and if $T = \max(t, \lambda)$, then the function $\delta_{t,\tau,\lambda}(x)$ tends to the soft thresholding function $\text{sgn}(x)(|x| - T)_+$ when τ tends to $+\infty$.

Now, when τ tends to infinity, $\delta_{0,\tau,\lambda}(x)$ tends to $\delta_{0,\infty,\lambda}(x)$, which is a hard thresholding function defined by:

$$\delta_{0,\infty,\lambda}(x) = \begin{cases} x \mathbb{1}_{\{|x|>\lambda\}} & \text{if } x \in \mathbb{R} \setminus \{-\lambda, \lambda\}, \\ \pm \lambda/2 & \text{if } x = \pm \lambda, \end{cases} \quad (24)$$

where $\mathbb{1}_\Delta$ is the indicator function of a given set $\Delta \subset \mathbb{R}$: $\mathbb{1}_\Delta(x) = 1$ if $x \in \Delta$; $\mathbb{1}_\Delta(x) = 0$ if $x \in \mathbb{R} \setminus \Delta$.

Note that $\delta_{0,\infty,\lambda}$ sets a coefficient with amplitude λ to half of its value and so, minimizes the local variation around λ , since $\lim_{x \rightarrow \lambda^+} \delta_{0,\infty,\lambda}(x) - 2\delta_{0,\infty,\lambda}(\lambda) + \lim_{x \rightarrow \lambda^-} \delta_{0,\infty,\lambda}(x) = 0$.

B Re-parameterization for the class of SSBS functions

Let $t \geq 0$ and $\lambda > t$. The sub-class of SSBS functions generated by varying τ admits two fixed points with non-null ordinates: in Cartesian-coordinates, $A = (\lambda, [\lambda - t]/2)$ and $A' = (-\lambda, -[\lambda - t]/2)$ belong to the curve of function $\delta_{t,\tau,\lambda}$ for every $\tau > 0$. Indeed, according to Eq. (30), we have $\delta_{t,\tau,\lambda}(\pm\lambda) = \pm[\lambda - t]/2$ for any $\tau > 0$. It follows that τ parameterizes the curvature of the arc of the SSBS function in the interval $]t, \lambda[$. This curvature directly relates to the attenuation degree we want to apply to data whose amplitudes belong to the interval $]t, \lambda[$.

Let C be the intersection between the abscissa axis and the tangent at point A to the curve of the SSBS function. The equation of this tangent is $y = 0.25(2 + \tau(\lambda - t))(x - \lambda) + 0.5(\lambda - t)$. The coordinates of point C are $C = ((2t + \tau\lambda(\lambda - t))/(2 + \tau(\lambda - t)), 0)$. We can easily control the arc \widehat{OA} curvature via the angle, denoted by θ , between vector \overrightarrow{OA} , which is fixed, and vector \overrightarrow{CA} , which is carried by the tangent to the curve of $\delta_{t,\tau,\lambda}$ at point A . The larger θ , the stronger the attenuation of the coefficients with amplitudes in $]t, \lambda[$. For fixed t and λ , the relation between angle θ and parameter τ is $\cos\theta = \overrightarrow{OA} \cdot \overrightarrow{CA} / (\|\overrightarrow{OA}\| \cdot \|\overrightarrow{CA}\|)$:

$$\cos\theta = \frac{10\lambda - 2t + \tau(\lambda - t)^2}{\sqrt{4\lambda^2 + [\lambda - t]^2} \sqrt{20 + 4\tau(\lambda - t) + \tau^2(\lambda - t)^2}}. \quad (25)$$

It easily follows from Eq. (25) that

$$0 < \theta < \arccos\left((\lambda - t) / \sqrt{4\lambda^2 + (\lambda - t)^2}\right).$$

When $\theta = \arccos((\lambda - t) / \sqrt{4\lambda^2 + (\lambda - t)^2})$, then $\tau = +\infty$, and $\delta_{t,\tau,\lambda}$ is the hard thresholding function of Eq. (24).

From Eq. (25), we derive that τ can be written as a function of t , θ and λ as follows:

$$\tau(t, \theta, \lambda) = \frac{1}{\lambda - t} \left(-\frac{\Gamma(t, \theta, \lambda)}{\Lambda(t, \theta, \lambda)} + \sqrt{\frac{\Gamma^2(t, \theta, \lambda)}{\Lambda^2(t, \theta, \lambda)} - \frac{20Y(t, \theta, \lambda) - 4(5\lambda - t)^2}{\Lambda(t, \theta, \lambda)}} \right) \quad (26)$$

where

$$\Upsilon(t, \theta, \lambda) = (\lambda^2 + (\lambda - t)^2) \cos^2 \theta,$$

$$\Gamma(t, \theta, \lambda) = 2(\Upsilon(t, \theta, \lambda) - (\lambda - t)(5\lambda - t)),$$

and

$$\Lambda(t, \theta, \lambda) = \Upsilon(t, \theta, \lambda) - (\lambda - t)^2.$$

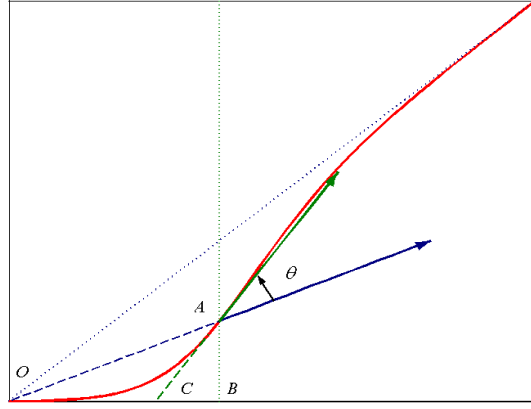


Figure 15: Graph of $\delta_{t,\tau,\lambda}$ in the positive half plan ($t=0$).

When no attenuation is required for large data, we are concerned by the particular case $t = 0$ (the SSBS shape is that of figure 15) and if we put $\delta_{\tau,\lambda} = \delta_{0,\tau,\lambda}$, Eqs. (2), (25) and (26) simplify:

$$\delta_{\tau,\lambda}(x) = \frac{x}{1 + e^{-\tau(|x|-\lambda)}}, \quad (30)$$

$$\cos \theta = \frac{10 + \tau \lambda}{\sqrt{5(20 + 4\tau \lambda + \tau^2 \lambda^2)}}, \quad (31)$$

with $0 < \theta < \arccos(\sqrt{5}/5)$, and

$$\tau(\theta, \lambda) = \frac{10}{\lambda} \frac{\sin^2 \theta + 2 \sin \theta \cos \theta}{5 \cos^2 \theta - 1}. \quad (32)$$